



# A Taxicab geometry quantification system to evaluate the performance of in silico methods: a case study on adenosine receptors ligands

Kamil J. Kuder<sup>1</sup> · Ilona Michalik<sup>1</sup> · Katarzyna Kieć-Kononowicz<sup>1</sup> · Peter Kolb<sup>2</sup>

Received: 19 June 2019 / Accepted: 14 February 2020 / Published online: 28 February 2020  
 © The Author(s) 2020

## Abstract

Among still comparatively few G protein-coupled receptors, the adenosine A<sub>2A</sub> receptor has been co-crystallized with several ligands, agonists as well as antagonists. It can thus serve as a template with a well-described orthosteric ligand binding region for adenosine receptors. As not all subtypes have been crystallized yet, and in order to investigate the usability of homology models in this context, multiple adenosine A<sub>1</sub> receptor (A<sub>1</sub>AR) homology models had been previously obtained and a library of lead-like compounds had been docked. As a result, a number of potent and one selective ligand toward the intended target have been identified. However, in in vitro experimental verification studies, many ligands also bound to the A<sub>2A</sub>AR and the A<sub>3</sub>AR subtypes. In this work we asked the question whether a classification of the ligands according to their selectivity was possible based on docking scores. Therefore, we built an A<sub>3</sub>AR homology model and docked all previously found ligands to all three receptor subtypes. As a metric, we employed an in vitro/in silico selectivity ranking system based on taxicab geometry and obtained a classification model with reasonable separation. In the next step, the method was validated with an external library of, selective ligands with similarly good performance. This classification system might also be useful in further screens.

**Keywords** Adenosine receptors · A<sub>1</sub>AR · A<sub>2A</sub>AR · A<sub>3</sub>AR GPCR · Homology modeling · Taxicab geometry · CBD

## Abbreviations

A <sub>1</sub> AR	Adenosine A <sub>1</sub> receptor
A <sub>2A</sub> AR	Adenosine A <sub>2A</sub> receptor
A <sub>3</sub> AR	Adenosine A <sub>3</sub> receptor
AR(s)	Adenosine receptor(s)
ASN	Asparagine
AUC	Area under the curve

BLAST	Basic local alignment search tool
bpc	Blocks per compound
CBD	City Block Distance
CNS	Central nervous system
CR	CBD calculation run
DOPE	Discrete optimized protein energy
DVR	Different partition CBD validation run
ECL	Extracellular loop
GPCRs	G protein-coupled receptors
ICL	Intracellular loop
PDB	Protein Data Bank/protein file format
PHE	Phenylalanine
QMEAN	Qualitative model energy analysis
RMS	Root mean square
ROC	Receiver-operator characteristic
SBDD	Structure-based drug design
TRP	Tryptophan
VR	CBD validation run
VS	Virtual screening

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10822-020-00301-5>) contains supplementary material, which is available to authorized users.

✉ Kamil J. Kuder  
[kamil.kuder@uj.edu.pl](mailto:kamil.kuder@uj.edu.pl)

✉ Peter Kolb  
[peter.kolb@uni-marburg.de](mailto:peter.kolb@uni-marburg.de)

<sup>1</sup> Department of Technology and Biotechnology of Drugs, Jagiellonian University Medical College, Medyczna 9, 30-688 Kraków, Poland

<sup>2</sup> Department of Pharmaceutical Chemistry, Phillips-University, Marbacher Weg 6, 35037 Marburg, Germany

## Introduction

A particular focus of rational drug design is the selectivity of novel ligands, with the aim to reduce possible side effects. The computational prediction of binding patterns of small molecules against multiple proteins would thus be of considerable interest. An *in silico* method that has worked particularly well with G protein-coupled receptors (GPCRs) is docking [1]. With respect to binding patterns, docking of a set of either newly designed ligands or virtual screening database compounds to various subtypes of a proposed biological target might narrow the group of potential ligands to those that exclusively interact with the intended protein(s).

GPCRs cover ~ 3% of the human proteome and represent the largest superfamily of membrane receptors. Built of seven transmembrane helices, they mediate signals from the out- to the inside of cells by sensing different agents. Binding of these agents leads to conformational changes and intracellular signaling cascades. Thus, GPCRs play a crucial role, either directly or indirectly, in the treatment of various pathophysiological states, evidenced by the fact that they are the targets of 30–50% of marketed drugs [2]. On the other hand, although the numbers have been rapidly increasing, only 62 members of the large family of GPCRs have been revealed as crystal structures up to date [3, 4]. Based on the fact that the transmembrane region of all GPCRs is well conserved, and knowing that most of class A GPCRs' ligand binding cavities are open toward the extracellular region [5], homology modeling provides a useful tool for structure-based ligand design. However, the accuracy of the models can be limited, mostly in the area of extra- and intracellular loops. The reason for that can be sought in highly variable loop sequences often corresponding to unaligned regions in sequence alignments, as well as their location at the solvent-exposed surface of proteins that result in higher conformational flexibility [6].

Adenosine is an important regulator for homeostasis of the brain, heart, kidney and other organs. It interacts with four different GPCRs classified as A<sub>1</sub>, A<sub>2A</sub>, A<sub>2B</sub> and A<sub>3</sub> subtypes. Selective interaction with adenosine receptor (AR) subtypes offers very broad therapeutic potential, including CNS disorders, regulation of electrophysiological properties of the heart, immune system and inflammatory diseases, cell growth, asthma, kidney failure and ischemic injuries [7]. Adenosine receptors' ligands are currently being developed as promising agents for CNS disorders (Parkinson's, Alzheimer's, epilepsy, ischemia) [8]. Also, the adenosine A<sub>2A</sub> receptor has been co-crystallized with several ligands, agonists as well as antagonists, and serves as a model AR with a well-defined orthosteric ligand binding region.

In order to investigate the usability of homology models for SBDD, multiple adenosine A<sub>1</sub> receptor (A<sub>1</sub>AR)

homology models have been previously obtained and a library of lead-like compounds has been docked [9]. As a result, a number of potent and a few selective ligands toward the intended target were found. However, in *in vitro* experimental verification studies many ligands also turned out to bind to A<sub>2A</sub>AR and A<sub>3</sub>AR. Therefore, the aim of this work was to build an A<sub>3</sub>AR homology model, generated on the basis of the evolutionarily closest homologous templates, and dock all previously used ligands to all three receptor subtypes. We wanted to investigate whether the experimentally obtained binding profiles [9] can be reproduced *in silico*, as well as to see how, instead of looking at individual compounds, the set of compounds is predicted within a given campaign. Although A<sub>1</sub>AR crystal structures have been published recently [10–12], in this study we used a series of homology models for the sake of consistency with our previous study. Likewise, an approach to construct a reliable *in silico/in vitro* correlation quantification system has been undertaken and its usability has been validated with an external library of highly selective ligands.

## Materials and methods

### Homology modeling

In order to find the most suitable protein template for the A<sub>3</sub>AR receptor model, its sequence was obtained from UniProt database [13] (sp\_P33765) and used for a BLAST search using two online tools: SwissModel [14] and Protein-Blast (NCBI) [15]. In both cases, the default search modes to find the most similar PDB crystal structures were used. After comparison of the results, three templates were chosen: 3EML (2.60 Å, 39.86% identity) [16], 2YDV (2.60 Å, 42.6% identity) [17] and 3VG9 (2.70 Å, 43.34% identity) [18]. Template proteins were chosen according to their highest crystallographic resolution (as well as crystal structures availability at the time) among two independent BLAST search hits, in order to increase the chances to obtain a reliable model.

Protein structures were pre-processed using PyMOL [19]: ligands, co-crystallization agents (2YDV, 3EML), the lysozyme insertion instead of ICL3 (3EML) were removed. Protein sequences obtained in this way were aligned using the PROMALS 3D [20] online tool. The resulting alignment, after visual inspection (position of transmembrane domains, possible disulfide bridges) was used as an input for MODELLER [21, 22].

Each of the 10 output models was then aligned to the 3EML crystal structure and carefully inspected visually using UCSF Chimera [23]. In particular, the orientation of the side chain of ASN250<sup>6.55</sup> (superscript numbers denote

Ballesteros–Weinstein numbers [24]) and other binding pocket amino acids was investigated and their possible, acceptable rotamers (according to the Dunbrack library [25]) were ascertained. Reasoning was supported by means of mutagenesis data [3, 4]. Similarly, we inspected the trans-membrane domains to avoid gaps, obvious steric clashes, unnatural side chain amino acid folding, as well as a preservation of the disulfide bonds between CYS83<sup>3.25</sup>–CYS166<sup>4.50</sup>.

### Known ligand database preparation

The next step was to test the enrichment of ligands over non-binders in the orthosteric binding pockets of the selected models. For this purpose, two sets of ligands were obtained from the ChEMBL database [26]. The “ligands” set consisted of approx. 1500 molecules described as A<sub>3</sub>AR ligands with a  $K_i \leq 100$  nM. Second, the decoy set consisted of approx. 800 molecules, tested against the A<sub>3</sub>AR and described as inactive for this target. Structures of both sets of ligands were obtained from the ZINC database [27] by searching for corresponding ZINC IDs for all of the ligands extracted from ChEMBL. High quality 3D conformer ensembles of both sets were obtained using the OMEGA module [28, 29] of the OEDocking software package (maximum number of conformers = 100; RMS = 0.5).

### Model refinement

The final A<sub>3</sub>AR homology model used in this study was obtained through the refinement process, using three different, consecutive strategies.

#### Strategy 1

As a reference ligand for docking, co-crystallized within the 3EML structure, the ligand ZM241385 was placed in the A<sub>3</sub>AR receptor models after their alignment to the 3EML structure, making sure that the hydrogen bonds with ASN250<sup>6.55</sup> were formed. Two sets of ligands were then docked to the prepared receptor homology models using the HYBRID module (one pose per ligand, max. hitlist size 500 molecules), implemented in the OEDocking Software [30–33]. After docking, the top 500 poses were inspected visually, and receiver operator characteristic (ROC) curves were generated along with calculations of the area under the curve (AUC), using an in-house script.

Docked ligands were minimized using the SZYBKI module (OEDocking) [34] and the homology model of the protein (model 1) was minimized (with ligand present in the binding site) using CHARMM [35]. As several unfavorable energy poses and similar docking behavior was observed

for the set of tested ligands, another modeling approach was then undertaken.

#### Strategy 2

Due to the fact that ZM241385 appears to be inactive towards A<sub>3</sub>AR and therefore might unduly bias the shape of the binding pocket during modeling, in a second round of modeling the previously identified ligand ZINC12533962, which is potent and selective towards A<sub>3</sub>AR (A<sub>3</sub>AR  $K_i = 40$  nM) was placed manually in the crystal structure of 3EML instead, retaining similar ligand-receptor interactions. As the conformation of the ligand, the previously obtained pose from docking to the A<sub>1</sub>AR [9] was used. The protein conformation prepared this way served as a template using the same input alignment for MODELLER as for model 1, excluding the 2YDV and 3VG9 X-ray structures and including the ligand and its position during modeling.

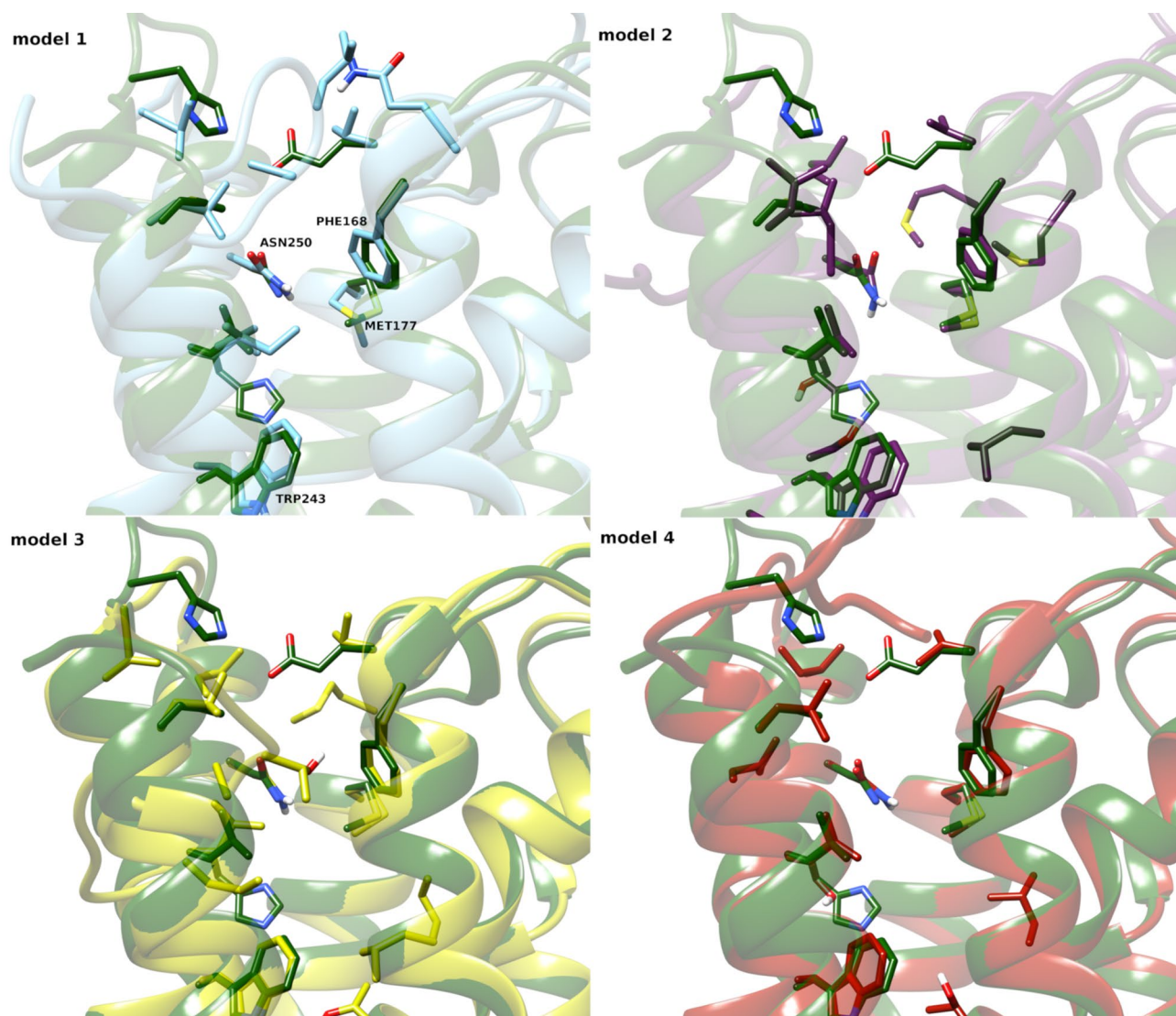
From the output of ten models, the best scoring one (according to Modeller’s scoring functions; model 2) was chosen for evaluation.

#### Strategy 3

Due to possible steric clashes between the ligand and TRP243<sup>6.48</sup> in the bottom part as well as PHE168<sup>4.52</sup> (ECL2) at the top of the binding pocket in model 2, again the position of ZINC12533962 in 3EML structure was corrected manually and the thus prepared receptor served again as a template for MODELLER. The resulting model (model 3) showed also potential steric clashes between the ligand and PHE168<sup>4.52</sup>, thus the ligand position in the template protein was again corrected, and the protein was remodeled. To the output model (model 4), after visual inspection, the set of actives and decoys was docked. This model (4), without further refinement and minimization, was chosen for all further docking studies. ECL2 was not remodeled, as it aligned well with the reference 3EML structure. Binding pockets of all four homology models obtained are presented in Fig. 1.

### Docking

In order to see whether the experimentally obtained binding profiles can be reproduced in silico, the next step involved docking of the previously described set of 39 ligands (test set) [9] to all three receptor subtypes (four A<sub>1</sub>AR homology models [9], the crystal structure of 3EML for A<sub>2A</sub>AR and the A<sub>3</sub>AR homology model). All ligands were prepared according to the same procedure described herein for the “binders/decoys” sets and docked to all receptors using the HYBRID module. As HYBRID docks multiconformer molecules into receptor-ligand complexes using an exhaustive search that systematically samples rotations and translations of each



**Fig. 1** The four obtained A<sub>3</sub>AR homology models (1–4, detailed information can be found in Model refinement section; left to right, upper and lower panel). Binding pocket residues were depicted as

thick sticks (only labeled in model 1 panel). The X-ray crystallographic structure of the template, A<sub>2A</sub>AR (3EML), is shown in dark green

conformer of the ligand within the active site (defined by the “bound” ligand), no docking grid/sphere was set beforehand. For all of the docked sets, the preservation of a hydrogen bond with ASN<sup>6.55</sup> as well as the orientation of the ligands in the binding pockets was inspected visually.

### In silico screening evaluation

For the quantification of the in silico/in vitro correlation, and, quite literally, to check how far from each other the results of those screenings are, an approach incorporating Taxicab geometry (City Block Distance, CBD) [36] and a traffic light system was utilized. Instead of the

usual distance in Euclidean geometry, Taxicab geometry defines a new metric in which the distance between two points ( $d_1$ ) is the sum of the absolute differences of their Cartesian coordinates ( $p$ ,  $q$ ).

$$d_1(p, q) = \|p - q\|_1 = \sum_{i=1}^n |p_i - q_i|$$

Among a variety of everyday life applications, CBD systems can also be used to assess the differences in discrete frequency distributions. In our study, instead of Cartesian coordinates, in vitro ( $v = K_i$ ) and in silico ( $s = \text{docking/rescore score}$ ) values were used for calculations.



$$d_1(v, s) = \|v - s\|_1 = \sum_{i=1}^n |v_i - s_i|$$

So as to organize the results of both, the in vitro and in silico screenings, the results were in a first instance classified empirically (Fig. 2), according to the following key:

- for in vitro values: green for  $K_i$  values in the range below 1000 nM, yellow for  $K_i$  values higher than 1000 nM, but still at a measurable level, red expresses no detectable binding.

- for in silico values: green expresses first 20% of the obtained docking score range, yellow next 20% of the obtained docking score range, red expresses the remaining 60% of docking score range (preliminary partitioning),

while the CBD values (0–2) were assigned to each color in the manner: 0 for red, 1 for yellow and 2 for green (CBD calculation run 1; *CR-1*). The division for in vitro data remained unchanged for further data development. However, it has been shown that the position of the ligand pose closest to the native pose is distributed rather randomly among

ZINC ID	<i>In vitro</i> % Inhibition*or $K_i$ [nM]			<i>In silico</i> - "CR-1" Chemgauss4 docking score			<i>In silico</i> - "CR-4" DrugScoreX		
	A <sub>1</sub>	A <sub>2A</sub>	A <sub>3</sub>	A <sub>1,4</sub>	A <sub>2A</sub>	A <sub>3</sub>	A <sub>1,4</sub>	A <sub>2A</sub>	A <sub>3</sub>
3445318	2 ± 2%	2360 ± 260	4860 ± 330	-7.717	-8.524	-10.515	-50.233	-100.036	-102.542
4059122	4 ± 4%	27 ± 6%	34 ± 2%	-9.661	-10.010	-8.914	-93.107	-74.122	-90.947
5238266	20.3 ± 20.3%	7.6 ± 4.3%	25.3 ± 6.4%	-9.166	-10.909	-9.375	-42.804	-96.027	-65.699
5809212	400 ± 60	740 ± 390	48 ± 7%	-6.683	-9.095	-5.740	-49.33	-87.148	-71.173
6498587	3340 ± 560	6660 ± 860	23 ± 3%	-12.771	-9.253	-9.513	-40.046	-91.874	-89.239
6647942	2 ± 2%	15 ± 4%	13 ± 3%	-9.106	-8.461	-8.913	-82.086	-100.897	-107.266
6823444	17 ± 9%	3310 ± 270	43 ± 3%	-8.264	-9.797	-9.109	-81.851	-96.674	-71.405
6844430	3.4 ± 3.4%	0.00%	11.1 ± 3.8%	-6.441	-9.428	-6.111	-50.554	-95.759	-49.713
6919900	28 ± 9%	3655 ± 870	2780 ± 920	-8.644	-9.898	-10.296	-63.619	-95.507	-93.195
6919944	10 ± 10%	10,900 ± 2200	3480 ± 1100	-11.992	-9.004	-10.504	-93.301	-100.401	-102.173
8653875	0.00%	2.1 ± 2.1%	10.4 ± 5%	-10.215	-9.179	-9.646	-60.361	-68.365	-82.529
8724684	10 ± 5%	18 ± 4%	46 ± 5%	-12.776	-10.139	-11.936	-82.444	-89.509	-91.467
8746576	28 ± 6%	20 ± 4%	38 ± 4%	-11.636	-13.397	-10.320	-36.719	-104.267	-90.706
12533962	33 ± 9%	6140 ± 1690	40 ± 6	-7.277	-8.093	-9.630	-69.986	-118.709	-97.569
13109814	3460 ± 420	11 ± 6%	35 ± 4%	-10.835	-9.327	-8.749	-101.031	-74.68	-85.743
16045161	3940 ± 390	1370 ± 470	3850 ± 590	-10.383	-7.913	-7.422	-63.108	-64.49	-78.982
18156418	10.9 ± 10.9%	8.8 ± 6.3%	26.8 ± 8.1%	-7.646	-10.908	-7.393	-82.613	-84.395	-80.455
18284742	8.9 ± 5.5%	8.1 ± 8.1%	20.7 ± 7.3%	-10.652	-11.690	-11.107	-65.896	-92.451	-98.47
18955638	19 ± 10%	6540 ± 1090	49 ± 1%	-8.993	-8.698	-7.843	-72.913	-78.355	-55.962
19849798	3430 ± 1030	2130 ± 720	1760 ± 210	-7.838	-6.266	-8.646	-63.238	-71.478	-75.701
20365432	33 ± 3%	2 ± 2%	19 ± 5%	-9.163	-9.112	-10.729	-5.076	0.46	-86.53
21934992	2930 ± 480	1450 ± 170	550 ± 70	-10.493	-10.526	-9.181	-93.719	-111.636	-106.076
28265812	13 ± 2%	36 ± 0.2%	13,400 ± 1900	-6.194	-7.477	-6.687	-72.856	-71.488	-62.928
28701281	45.00%	3560 ± 510	1520 ± 360	-8.181	-10.500	-9.715	-85.039	-55.599	-82.16
29256877	13 ± 9%	35 ± 3%	13,700 ± 2200	-8.831	-8.663	-7.516	-89.661	-93.384	-42.784
29354708	1220 ± 340	3780 ± 830	70 ± 20	-7.898	-8.357	-10.673	-101.401	-94.509	-104.07
29462316	980 ± 90	1340 ± 310	205 ± 30	-8.672	-9.069	-8.204	-80.587	-107.017	-91.341
29573785	9 ± 9%	37 ± 1%	9060 ± 1100	-11.211	-10.019	-12.100	-102.876	-68.215	-97.123
30654128	36.00%	9300 ± 700	42 ± 6%	-10.012	-4.807	-6.920	-45.14	23.755	-61.393
31402186	20 ± 4%	5 ± 3%	9330 ± 1800	-10.237	-10.123	-9.635	-73.847	-105.348	-111.032
31584575	12 ± 7%	12 ± 6%	19 ± 4%	-13.352	-7.851	-7.371	33.954	-93.334	-69.914
31875387	11 ± 1%	1 ± 1%	16 ± 7%	-8.594	-10.934	-8.133	-83.077	-94.293	-69.013
40488321	21 ± 2%	37 ± 4%	41 ± 2%	-11.284	-11.387	-10.533	-116.482	-108.804	-91.589
43348592	3 ± 2%	4 ± 4%	23 ± 6%	-8.295	-6.634	-7.956	-71.531	-79.09	-92.342
47260835	8 ± 4%	14 ± 3%	20 ± 5%	-6.257	-4.938	-6.162	-79.993	-34.856	-69.093
47260884	4 ± 4%	48 ± 3%	31 ± 6%	-9.028	-7.158	-5.896	-40.962	-63.131	-59.752
47260934	10 ± 8%	53 ± 3%	31 ± 2%	-7.801	-7.700	-6.585	-28.219	-80.974	-58.454
54203984	3 ± 3%	17 ± 5%	28 ± 2%	-10.746	-7.763	-6.605	-55.481	-63.983	-72.343
56482974	5 ± 4%	45 ± 5%	53 ± 3%	-8.564	-8.845	-9.162	-69.005	-72.088	-70.51

\*% Inhibition at 10 μM compound concentration

**Fig. 2** Distribution of CBD ranks for in vitro values (left panel), in silico, first docking run ("CR-1", middle panel) and in silico second docking run with following rescoring ("CR-4", right panel). Rank colors are assigned as stated in text. Detailed binding data can be found in [9]

all generated poses and ordered with respect to the docking score [37]. Hence, the docking results were rescored using DSX-Online and the color scheme was adapted as follows: green  $< -100$ ,  $-100 < \text{yellow} < -90$ , red  $> -90$  (Fig. 2; CR-2). The next step of proposed platform evaluation was redocking of the whole set of ligands to all four adenosine A<sub>1</sub> receptor models, using ZINC12533962 as a reference ligand instead of ZM241385, in order to obtain a fair comparison for the docking procedure. This has been done by overlaying the obtained A<sub>3</sub>AR homology model onto the backbone of the A<sub>1</sub>AR models and preserving the coordinates of ZM241385. Using the same data partitioning as in the preliminary calculations, a CBD value was calculated (CR-3). A rescoring procedure was incorporated as described above (CR-4). To eliminate potential boundary effects arising from the in vitro/in silico data partitioning, the system was changed to a binary distribution (CR-5–CR-8 for each previous run respectively) for either active or nonactive for the biological target (CBD = 1 for previous greens and yellows, 0 for reds) and recalculated (Fig. 4). Moreover, in order to determine the relative CBD value ( $\text{CBD}_{\text{rel}} = \text{CBD} / \text{CBD}_{\text{max}}$ ), the maximal possible CBD ( $\text{CBD}_{\text{max}}$ ) values for each distribution were calculated. A  $\text{CBD}_{\text{rel}}$  of less than 1 indicates better-than-random performance.

## Method validation

In order to test the usability and versatility of the described method, a library of 88 selective ligands previously described by Katritch et al. [38] was used. This ligand database was prepared according to the same procedure described herein for the “binders/decoys” sets and docked to all receptors using the HYBRID module. Likewise, the same data partition system as described in “In silico screening evaluation” subsection was applied. However, due to the high affinity of the ligands, the second system was incorporated:

- for in vitro values: green for  $K_i$  values in the range below 100 nM, yellow for  $K_i$  values between 100 and 1000 nM, red expresses over 1000 nM or no detectable binding

Detailed information on the used set along with partitioning systems incorporated can be found in the Supplementary Material.

## Results

### Homology modeling

The homology model of the A<sub>3</sub>AR developed for this study exhibited good quality, as characterized by the fact that

97.2% residues are in the favored region of the Ramachandran plot.

As MODELLER outcome, ten models (#mo1–#mo10; Modeller output) were constructed using the input alignment. Models were characterized by relatively high DOPE (Discrete Optimized Protein Energy, atomic distance-dependent statistical function), GA341 (describing reliability of a model, derived from statistical potentials) scores and molpdf values (molecular PDF, Modeller objective function—the sum of all restraints). The three best-scoring A<sub>3</sub>AR models (denoted: #mo3, #mo9, #mo2) were used for further evaluation strategies based on two main criteria: AUC of binders/non-binders docking evaluation and visual inspection of the docked ligands. The most convincing binding modes, as well as an AUC = 0.776 were observed for model #mo3, which was consequently chosen for further studies, and denoted as model 1.

Despite its high enrichment and acceptable binding modes of the “active” set of ligands, we have to note that possible steric clashes between ligands and amino acids were observed, as well as narrowing of the bottom part of binding pocket, indicating a low quality model that might lead to false results. Therefore another modeling approach (denoted as Strategy 2) was undertaken.

Model 2 (molpdf: 1823.76904, DOPE score –41,914.38281) was obtained as a result of strategy 2, by placing the highly potent and selective (A<sub>3</sub>AR  $K_i = 40$  nM) triazine-based molecule ZINC12533962, into the binding pocket instead of ZM241385, the ligand inactive for A<sub>3</sub>AR, but co-crystallized within the 3EML structure, used again as a template. Since ZINC12533962 was a hit derived from the previous study [9] its selection for purpose of this study was straightforward. Models 3 (molpdf: 3149.86035, DOPE score: –41,785.51562) and 4 (molpdf: 1593.80505, DOPE score: –41,855.65625; used further in docking/selectivity studies) were the results of strategy 3. This entailed modeling after manual corrections of the ZINC12533962 ligand in the template structure, in order to obtain models with high enrichment ratios and possibly no steric clashes. The final model 4 used for the docking studies was characterized by a QMEAN score of 0.512 [39, 40], and 97.2% residues in the favored region according to the Ramachandran plot [41]. For all of the models, the orientation of the side chains in the binding side and the preservation of disulfide bonds was also checked visually. Binding pockets of all four obtained homology models are presented in Fig. 1.

### Known ligand database docking and CBD system evaluation

As expected, most of the previously selected 39 test set ligands docked to the newly obtained A<sub>3</sub>AR homology model forming two hydrogen bonds with the key residue

ASN250<sup>6,55</sup> in the calculated poses. Also, for most in vitro active compounds,  $\pi$ - $\pi$  stacking interactions between their (hetero)aromatic rings and PHE168<sup>45,52</sup> were observed. Figure 3 depicts exemplary binding mode to the adenosine A<sub>3</sub> receptor homology model 4.

In fact, in terms of productivity, our adenosine A<sub>3</sub>AR model gave an average true positive rate of 27 correct in silico/in vitro matches, out of 39 docked structures. A similar tendency was observed for all four used adenosine A<sub>1</sub>AR models, in contrast to, interestingly, a slightly lower score obtained using the crystal structure of the adenosine A<sub>2A</sub> receptor. Also, exchange of the “original” ligand in the binding pocket of the A<sub>1</sub>AR models did not affect the overall productivity score. On the other hand, rescoring with DSX slightly increased the hit/non-hit ratio for the A<sub>1</sub>AR and A<sub>2A</sub>AR results.

A similar influence was observed for CBD system incorporated in this study for selectivity prediction purposes (CR-1 CBD = 61, avg = 1.56; CR-3 CBD = 56, avg = 1.44). Rescoring, by increasing the productivity of models, resulted in slightly decreasing the CBD value in the first docking procedure (CR-2 CBD = 60, avg = 1.53; CR-4 CBD = 58, avg = 1.49). By changing the computational data partitioning to a binary system (either active or inactive, “greens & reds”, CR-5–CR-8) we were able to obtain CBD values of 46 for the docking (CR-5), and 37 for the rescoring procedures (CR-8). The same tendency was observed when calculating CBD values for particular targets (e.g. 13 vs. 12 for A<sub>3</sub>AR, Fig. 4) and relative CBD values as well. With CBD<sub>max</sub> value

of 172 for ‘0–1–2’ and 117 for binary distributions at given in vitro data, the CBD<sub>rel</sub> values decreased from 0.36 (CR-1) to 0.32 (CR-8), proving the effectiveness of conducted calculations.

Overall, using the a priori data partitioning, starting from the average error of 1.56 blocks per compound (bpc), we were able to improve the CBD ranking method to an average error of 1.49 bpc. Binarization of the data partitioning allowed to decrease the distance even further, from average error of 1.18 bpc to 0.95 bpc.

### CBD system validation

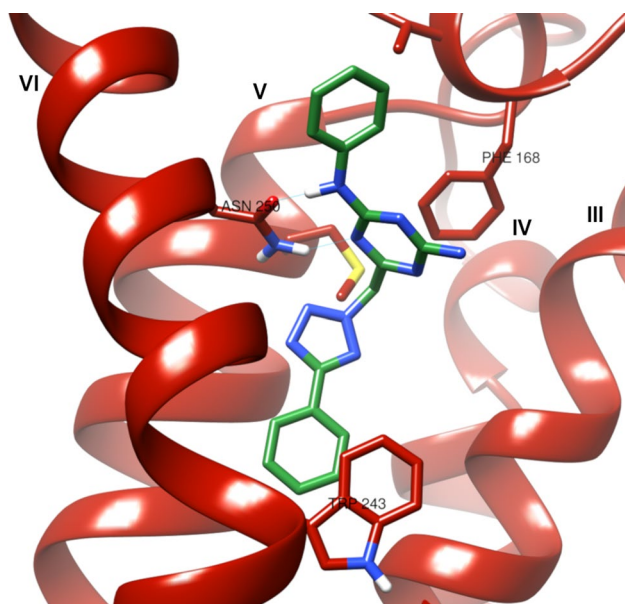
The usability and versatility of the described CBD-based method has been assessed also with a set of reference compounds with strong selectivity, described by Katritch [38]. While the obtained CBD results of these validation runs (“VR”, for details please refer to Supplementary Material) were higher than for the test set, they still bore correlation to the CR’s. Starting from VR-1 CBD resulted in avg. = 2.76 (when compared to 1.56 for test set), the method allowed for VR-8 CBD avg. = 1.17. While the absolute values are somewhat different, a strong correlation with an R value of 0.96 ( $R^2 = 0.92$ ) between CR and VR runs can be determined. A tendency similar to the one for the test set—decreasing CBD<sub>rel</sub> values for each run—was also observed. With CBD<sub>max</sub> value of 476 for ‘0–1–2’ and 264 for binary distributions at given in vitro data, the CBD<sub>rel</sub> values decreased from 0.51 (VR-1) to 0.39 (VR-8).

While the validation set expressed affinity for biological targets at much higher level than the test set, the sensitivity of method was increased, by exchanging the in vitro data partitioning system (Second Validation Run; SVR). With sensitivity set at  $K_i < 100$  nM for greens and yellows for  $100 \text{ nM} < K_i < 1000 \text{ nM}$ , CBD avg. obtained CBD values appeared slightly higher than those of VR’s (CR/SVR CBD  $R = 0.88$ ,  $R^2 = 0.77$ ). However, the trend of decreasing CBD<sub>rel</sub> values was maintained, and remained at the ~0.4 level.

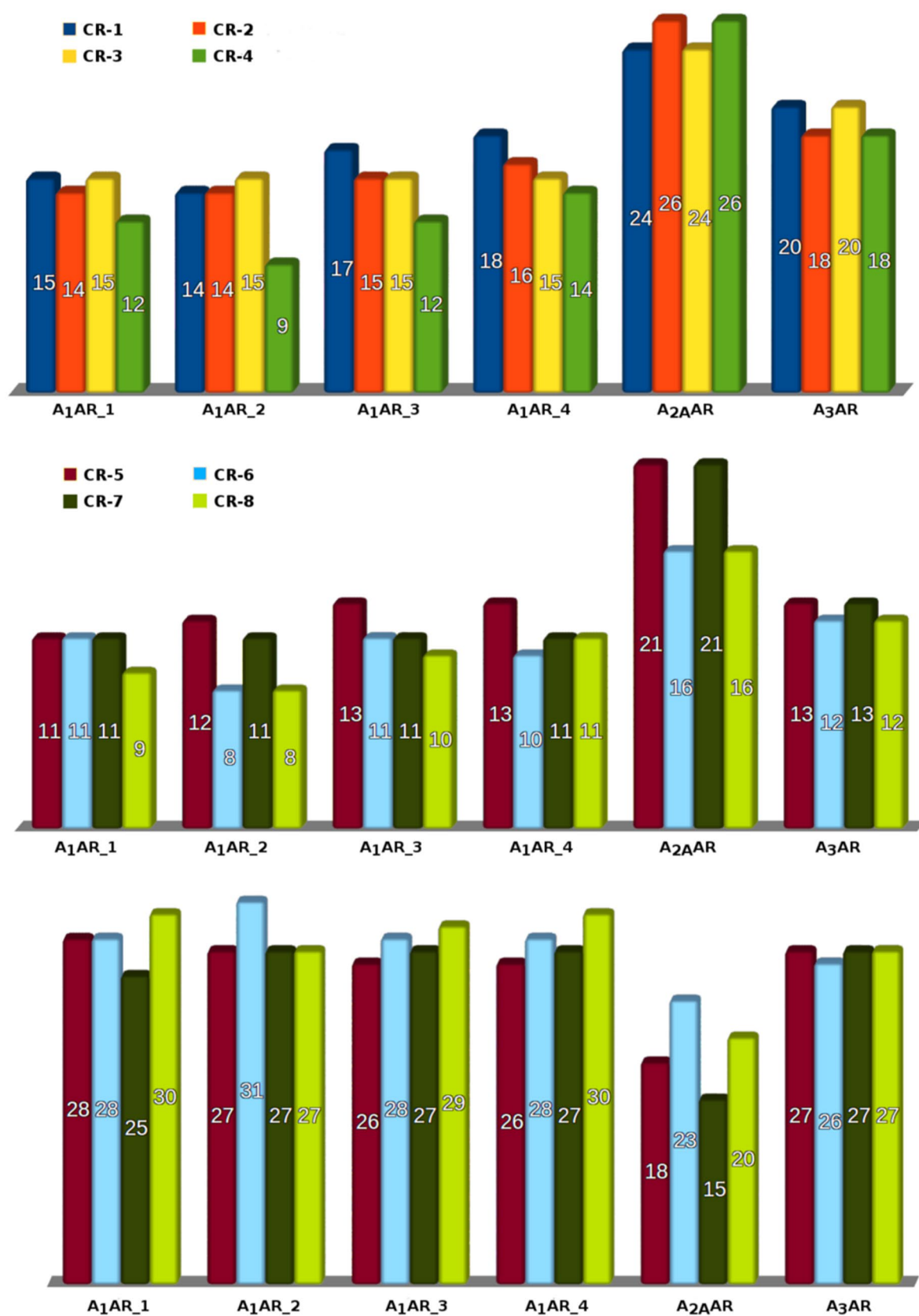
In light of these validation results, we think that the ranking system herein presented can be used as an in silico/in vitro correlation quantification system independently of ligand selectivity.

### Discussion

Predicting the subtype selectivity of ligands to GPCRs using in silico methods still remains a challenge for modelers [37] for a number of receptors share a high degree of structural similarity among their subtypes. Adenosine receptors are no different indeed—all AR subtypes share core interactions within their conserved residues. These interactions include



**Fig. 3** Calculated binding pose of ZINC 12533962 in adenosine A<sub>3</sub> homology receptor (model 4) binding pocket. Blue lines indicate hydrogen bonds formed with ASN250<sup>6,55</sup>. Latin numbers indicate helices, helix VII was removed for better viewing purposes





**Fig. 4** Upper panel: distribution of CBD values calculated in 0–1–2 mode for each receptor used; Middle panel: distribution of CBD values calculated in binary mode for each receptor; Lower panel: in vitro/in silico hit rate for all 6 used proteins (four A<sub>1</sub>AR models and one each for A<sub>2A</sub>AR and A<sub>3</sub>AR). Each column depicts one run according to the legend. For detailed information on proteins used, as well as “CR’s” please refer to “Materials and Methods”: Docking and In silico screening evaluation sections respectively

key strong hydrogen bonding with Asparagine in position 6.55, aromatic stacking with Phenylalanine in position 5.29 along with hydrophobic interactions with conserved Iso-leucine and Leucine side chains in positions 7.39 and 6.51, respectively. Nonetheless, small mutations in overall highly homologous structures serve as key selectivity determinants for AR subtypes. In case of A<sub>1</sub>AR, that has a close similarity to A<sub>2A</sub>AR and A<sub>2B</sub>AR, a difference is caused by four mutations in the periphery of the binding pocket (e.g. position 7.35) and relatively shorter ECL2, which results in the formation of an additional hydrophobic sub-pocket in the loop region, allowing ligands to be more mobile in the binding pocket [42]. On the other hand, A<sub>3</sub>AR is the most disparate among ARs, having 10 (out of 20) unique amino acids in the binding pocket. The key difference is believed to be caused by exclusive Glutamate to Valine replacement in position 5.30 which plays an important role in ligand binding to for the remaining ARs [38].

Therefore, finding a ligand that is recognized by just one binding pocket and refused by all other subtypes heavily relies on the quality of protein models used for the studies, as well as the docking procedure, and data analysis. Up to date, a number of virtual screening approaches to find such ligands for adenosine receptors have been undertaken. Just to mention the latest few: Rodriguez et al. [43] were able to identify 9 A<sub>2A</sub>AR ligands out of 20 predicted while scanning a library of 6.7 million compounds, however none of them activated the target receptor. Later studies from this group incorporating virtual libraries allowed for the identification of two ligands targeting A<sub>1</sub> and A<sub>3</sub>AR [44]. From the 63 structurally diverse ligands identified by VS by Tian et al. [45], 11 exhibited substantial activity against A<sub>2A</sub>AR in experimental tests, 2 of which with K<sub>i</sub> in nanomolar level and good A<sub>2A</sub>/A<sub>1</sub> selectivity. Last but not least, a non-typical, interesting incorporation of virtual screening in later stage of drug development, namely safety profiling, was described recently by Fan et al. [46].

Even though the selectivity of the compounds in the selected dataset was not extraordinary and we did not have the highest-affinity ligands known for the ARs in this set, for the sake of consistency with previous studies we used the same receptors and ligand sets. This also ensured data comparability. Moreover, we wanted to see how, instead of looking at individual compounds, the entire *set* of compounds is predicted—within a real-world campaign, one might not

have a validated set of high-affinity ligands. From this study two main results emerged. First, our A<sub>3</sub>AR model is a reliable one in the sense that it is capable of recognizing the active ligands with high confidence (AUC = 0.844). This was also proven by accurate predictions for 27 out of the 39 ligands used in this study. The overall hit ratio is comparable with the A<sub>1</sub>AR homology models used in this study. Yet, in comparison to the A<sub>1</sub>AR models used for screening, the docking procedure involved only one A<sub>3</sub>AR homology receptor in a rigid state. Taking into account the high GPCR flexibility and the fact that only one out of a number of possible receptor conformations was used to obtain a relatively good hit ratio, it might be concluded that the obtained model can be further used for screening of larger libraries of compounds.

Second, the employed in silico ranking method based on Taxicab Geometry proved to be a useful metric to evaluate the performance of the in silico methods. The idea was to estimate and visualize how far from “exact” the screening results would be. As this study was the continuation of a previously described one [9] we already had the binding data in hand. Therefore, dividing the data into active and inactive compounds was the first step of applying the method.

At this point one has to deal with the uncertainties associated with experimental data. For this reason, and also to avoid classifying e.g. two compounds with 999 nM and 1001 nM affinity, respectively, as “active” and “inactive”, we introduced the “buffer” category yellow. The same applies for molecules with a percentage of inhibition close to 50%. In the docking part of our study, we chose to impose the set delimiters by percentage of the entire set, thus avoiding the issue of category-crossing error bars on e.g. ‘greens’ and ‘yellows’. By using a percentage-based splitting and focusing on the correlation of the entire set of molecules, the performance and assignment of an individual compound does not influence the overall performance to a large extent. Of course, in a prospective setting this percentage-based separation into categories is not obvious. Users might be guided by the commonly achieved hit rates in such cases, or, alternatively, by the capacity available for experimental testing.

On the other hand, dividing the computational docking data a priori was challenging. How to divide the docking results without being biased by the biological test data? This was even amplified when incorporating the third category, “moderately active”, to both sets of data to make the estimation more precise. However, even when juggling and correcting the screening data division, followed by redocking as well as rescoring of all of the poses, we were able to only get CBD values quite far from ideal. Nonetheless, comparing these values with an in-house script generating multiple 0–1–2 distributions, we were able to ascertain that the results were better than random. Further changing the data division to binary, either active

or inactive, in fact yielded lower CBD values, resulting in an average error of less than 1 block per compound. Such a rank division seems satisfying, due to a high hit/no-hit ratio yield, clearly better than random, and should be accurate enough to be used also in a prospective setting. Again, using the same script generating multiple 0–1 distributions, results also appeared to be slightly better than random. The strength and effectiveness of these calculations were confirmed with relative CBD values calculated for each run, clearly proving their non-randomness.

As a further matter, for the purpose of method validation, we used a set of 88 highly affine and selective ligands. Albeit the obtained results appeared slightly higher than for the test set, they still prove the effectiveness of the method in both 0–1–2 and 0–1 distributions. The main reason for such performance might be the overall higher ligand affinity to the adenosine receptors. This issue was partly resolved by fine-tuning of the method, by increasing the sensitivity of in vitro data partitioning (eg. *VR-1* CBD avg. = 2.76 vs. *SVR-1* CBD avg. = 2.12). It also allowed us to show that such a sensitivity increase is not only possible, but also productive.

Nevertheless, docking to multiple receptors and ranking the data might be a challenge. First, four different homology models of one receptor (*A<sub>1</sub>AR*) generated only small differences in binding values, as the models only differ slightly from each other. Taking into account ranks from all four models seems like a sensible strategy, but care has to be taken not to bias the overall ranking just for this one target. Second, with each additional target and the set of screened ligands, the amount of data combinations increases, and ranks have to be divided carefully. Third, one must keep in mind that the assignment of docking ranks and scores to the different categories might still be influenced by the researcher deciding on the cutoffs.

Contemporary searches for novel GPCRs ligands rely heavily on docking of newly designed chemical compounds and virtual libraries to protein homology models and crystal structures. Still, the exact identification of hits and prediction of their selectivity profiles remains a challenge for computational chemists. The ranking system described herein might find its use in the search for selective compounds, but also those that are designed to act at more than one target. Despite its simplicity, it helps to condense a rather complex comparison into single numbers and cutoffs for classifiers.

**Acknowledgements** Studies presented herein were financed on the basis of an STSM, Grant No ECOST-STSM-CM1207-020215-054712 of the COST Action “GLISTEN”, as well as Jagiellonian University Medical College Statutory Funding for the Development of Science: N42/DBS/000041; N42/DBS/000039 and National Science Center, Poland granted on the basis of decision No. DEC-2018/02/X/NZ7/00584. Authors acknowledge the support of COST Action CM1207—GLISTEN.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Kufareva I, Katritch V, Participants of GPCR Dock (2013) G, Stevens RC, Abagyan R (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Struct Des* 22:1120–1139
- Garland SL (2013) Are GPCRs still a source of new targets? *J Biomol Screen* 8:947–966
- Munk C, Isberg V, Mordalski S, Harpsøe K, Rataj K, Hauser AS, Kolb P, Bojarski AJ, Vriend G, Gloriam DE (2016) GPCRdb: the G protein-coupled receptor database—an introduction. *Br J Pharmacol* 173:2195–2207
- Isberg V, Mordalski S, Munk C, Rataj K, Harpsøe K, Hauser AS, Vroeling B, Bojarski AJ, Vriend G, Gloriam DE (2016) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* 44(D1):275–279
- Jacobson KA, Costanzi S (2012) New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. *Mol Pharmacol* 82:361–371
- Fiser A (2010) Template-based protein structure modeling. *Methods Mol Biol* 673:73–94
- Jacobson KA, Müller CE (2016) Medicinal chemistry of adenosine, P2Y and P2X receptors. *Neuropharmacology* 104:31–49
- Stockwell J, Jakova S, Cayabyab FS (2017) Adenosine A1 and A2A receptors in the brain: current research and their role in neurodegeneration. *Molecules* 22:676
- Kolb P, Phan K, Gao Z-G, Marko AC, Sali A, Jacobson KA (2012) Limits of ligand selectivity from docking to models: in silico screening for A1 adenosine receptor antagonists. *PLoS ONE* 7:e49910
- Glukhova A, Thal DM, Nguyenm AT, Vevvhio EA, Jörg M, Scammells PJ, May LT, Sexton PM, Christopoulos A (2017) Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* 168:867–877
- Cheng RKY, Segala E, Robertson N, Defforian F, Dore AS, Errey JC, Fiez-Vandal C, Marshall FH, Cooke RM (2017) Structures of human A1 and A2A adenosine receptors with Xanthines reveal determinants of selectivity. *Structure* 25:1275–1285
- Draper-Joyce CJ, Khoshouei M, Thal DM, Liang YL, Nguyen ATN, Furness SGB, Venugopal H, Baltos JA, Plitzko JM, Danev R, Baumeister W, May LT, Wootten D, Sexton PM, Glukhova A, Christopoulos A (2018) Structure of the adenosine-bound human adenosine A1 receptor-Gi complex. *Nature* 558:559–563
- UniProt, <https://www.uniprot.org/>. Accessed 14 Jun 2019
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42(W1):W252–W258
- Protein BLAST. <https://blast.ncbi.nlm.nih.gov/>. Accessed 14 Jun 2019

16. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien YET, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 Å crystal structure of a human A<sub>2A</sub> adenosine receptor bound to an antagonist. *Science* 322:1211–1217
17. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A<sub>2A</sub> receptor structures reveal common features of GPCR activation. *Nature* 474:521–525
18. Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomur N, Cameron AD (2012) G-Protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* 482:237–240
19. The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC
20. Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36:2295–2300
21. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A (2007) Comparative protein structure modeling using MODELLER. In *Current Protocols in Protein Science*. Wiley, Hoboken
22. Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
24. Ballesteros JA, Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G Protein-coupled receptors. *Methods Neurosci* 25:366–428
25. Dunbrack RL (2002) Rotamer libraries in the 21st century. *Curr Opin Struct Biol* 12:431–440
26. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Krü FA, Light Y, Mak L, Mcglinchey S, Nowotka M, Papadatos G, Santos R, Overington JP (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res* 42:D1083–1090
27. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 42:177–182
28. OMEGA 3.0.0.1: OpenEye Scientific Software, Santa Fe, NM. <https://www.eyesopen.com>
29. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* 50:572–584
30. OEDOCKING 3.2.0.2; OpenEye Scientific Software, Santa Fe, NM. <https://www.eyesopen.com>
31. Kelley BP, Brown SP, Warren GL, Muchmore SW (2015) POSIT: flexible shape-guided docking for pose prediction. *J Chem Inf Model* 55:1771–1780
32. McGann M (2011) FRED pose prediction and virtual screening accuracy. *J Chem Inf Model* 51:578–596
33. McGann M (2012) FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des* 26:897–906
34. SZYBK. OpenEye Scientific Software, Santa Fe, NM. <https://www.eyesopen.com>
35. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217
36. Krause EF (1986) *Taxicab geometry: an adventure in Non-Euclidean geometry*. Dover Publications, New York
37. Plewczynski D, Łażniewski M, Augustyniak R, Ginalski K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J Comput Chem* 32:742–755
38. Katritch V, Kufareva I, Abagyan R (2011) Structure based prediction of subtype-selectivity for adenosine receptor antagonists. *Neuropharm* 60:108–115
39. Benkert P, Kü Nzli M, Schwede T (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res* 37:W510–W514
40. Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. *Proteins Struct Funct Bioinform* 71:261–277
41. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C $\alpha$  geometry:  $\Phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins Struct Funct Bioinform* 50:437–450
42. Mattedi G, Deflorian F, Mason JS, de Graaf C, Gervasio FL (2019) Understanding ligand binding selectivity in a prototypical GPCR family. *J Chem Inf Model* 59:2830–2836
43. Rodriguez D, Gao Z-G, Moss SM, Jacobson KA, Carlsson J (2015) Molecular docking screening using agonist-bound gpcr structures: probing the A<sub>2A</sub> adenosine receptor. *J Chem Inf Model* 55:550–563
44. Rodriguez D, Chakraborty S, Warnick E, Crane E, Gao A-G, O'Connor R, Jacobson KA, Carlsson J (2016) Structure-based screening of uncharted chemical space for atypical adenosine receptor agonists. *ACS Chem Biol* 11:2763–2772
45. Tian S, Wang X, Li L, Zhang X, Li Y, Zhu Y, Hou T, Zhen X (2019) Discovery of novel and selective adenosine A<sub>2A</sub> receptor antagonists for treating Parkinson's disease through comparative structure-based Virtual screening. *J Chem Inf Model* 57:1474–1484
46. Fan F, Warshaviak DT, Hamadeh HK, Dunn RT II (2019) The integration of pharmacophore-based 3D QSAR modeling and virtual screening in safety profiling: a case study to identify antagonistic activities against adenosine receptor, A<sub>2A</sub>, using 1,897 known drugs. *PLoS* 14:e0204378

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.